

Gatekeeping, Self-Selection, and Utilization of Curative and Preventive Health Care Services*

Partha Deb

Department of Economics

Indiana University-Purdue University Indianapolis

Hunter College, City University of New York

Pravin K. Trivedi†

Department of Economics

Indiana University

March 2003

Abstract

We examine how the gatekeeper role of health maintenance organizations (HMO) and other managed care health insurance plans impacts five types of curative services and five preventive care services. We accommodate endogeneity of plan choice by incorporating latent factors to control for possible selection, and estimate these models using accelerated simulated likelihood methods. Using the nationally representative 1996 Medical Expenditure Panel Survey, we find robust evidence of selection bias. After accounting for selection, we find that HMOs with gatekeeping features encourage the use of physician, outpatient hospital and emergency room care and the use of most preventive services.

Keywords: Endogenous treatment; factor loadings; managed care; health care utilization

JEL Codes: C35, C51, I11

*We gratefully acknowledge research support from the Agency for Health Research and Quality (R01 HS10904-02). We wish to thank David Zimmer for excellent research assistance and Randy Ellis, Ann Holmes and Tom McGuire for very helpful suggestions.

†Corresponding author: Department of Economics, Indiana University, Wylie Hall, Bloomington, Indiana 47405, USA. Email: trivedi@indiana.edu, Phone: 812-855-3567, Fax: 812-855-3736.

1. Introduction

Managed care plans seek to provide cost-effective care by using a variety of financial and nonfinancial tools to manage care. These include selecting a network of providers, deemphasizing specialist care while relying on primary care, using primary care physicians as gatekeepers to specialists, using financial incentives to encourage cost containment, and so forth. Among these, plans with gatekeepers have the most direct provider-side control on the use of services by consumers. Gatekeeping is an identifying feature of health maintenance organizations (HMO) and exists in many preferred provider organizations (PPO) and point of service plans (POS). Although gatekeeping was hailed as the solution to the problem of moral hazard in the early years, it has been recently demonized in the popular press and in public opinion for being too restrictive. Nevertheless, it continues to be an important feature of managed care plans with tight utilization controls.

In this paper we develop a model to estimate the causal effect of gatekeeping on utilization of curative and preventive health care services. Gatekeeping is defined by categorizing plans into HMOs (which always feature gatekeeping), other managed care plans that feature gatekeeping and plans without gatekeeping. We construct a joint distribution of endogenous health insurance choice and utilization using a latent factors specification. Latent factors are incorporated into the insurance and utilization equations to allow for idiosyncratic influences on insurance plan choice to affect utilization, thus enabling us to make a distinction between selection on unobservables and selection on observables (Heckman and Vytlacil, 2001). We interpret these idiosyncratic influences as unobserved heterogeneity. The model captures heterogeneity in the utilization response to insurance plans, which is known to be an important feature of impact of job training programs (Heckman, Smith and Clements, 1997).

Both treatment and outcome processes are non-normal and nonlinear (multinomial, count, discrete) and treatment is endogenous. In such models, linear instrumental variables methods are generally inappropriate and nonlinear instrumental variables are also either inappropriate or do not work very well. Consequently,

we apply maximum simulated likelihood (MSL) techniques to estimate the parameters of our models. Simulation is used to evaluate integral expressions in the likelihood function of the model as no closed form solutions exist.

Studies have shown that HMOs are associated with lower hospitalization rates, reduced lengths of stay, either unchanged or more office visits, and greater use of preventive services (for reviews, see Glied, 2000; Miller and Luft, 1994; Cutler and Zeckhauser, 2000). But, conclusions continue to be unsettled due to institutional changes in the 1990's (Gabel, 1997) combined with statistical and generalizability problems in the literature. A major statistical source of ambiguity in previous analyses of health care service utilization is due to the relative neglect of the problem of endogeneity bias caused by self-selection into health insurance plans. Economic models of the choice of health insurance and medical care utilization provide strong a priori justification for treating insurance choice as endogenous and jointly determined along with health care utilization (Cameron, et al., 1988; Gilleskie, 1998). Goldman (1995) and Mello, et al. (2002) explicitly address the issue of self-selection on unobservables and find that selection bias in effects of health insurance choice on utilization can be substantial. Yet, empirical research in health services either ignores the role of self-selection or acknowledge the possibility without solutions (Christensen and Shinogle, 1997; Tu, Kemper and Wong, 1999). Hence, some of the more authoritative findings on the role of managed care cited in the literature continue to be based on the RAND Health Insurance Experiment (Newhouse, et al., 1993). However, given the age of this experiment and the major changes in health care technology and delivery systems, including dramatic evolution in the structure of managed care plans and HMOs, the continued relevance and validity of these findings in the current environment may be questioned.

The remainder of this paper is organized as follows. Section 2 presents a theoretical framework that describes the mechanisms by which selection into plans operates and how that affects health care utilization. Section 3 describes the econometric framework. The data are described in Section 4. Section 5 presents

and discusses the empirical results. Section 6 concludes.

2. Theoretical Framework

We begin by presenting a theoretical framework based on the consumer as the economic agent. Our emphasis on the consumer as the agent is in keeping with our econometric model and much prior work. However, selection can arise in other ways, for example in models that are based on the economic behavior of health plans (Frank, et al., 2000; Cao and McGuire, 2002). Following our theoretical model, we describe their main features.

Our starting point is provided by a simplified version of the model given in Cameron, et al. (1988). In this model a consumer maximizes a utility function, defined by $U[C, H(\mathbf{y}, s|\mathbf{A}, d_j)]$, where C denotes consumption of other goods, H denotes a “health production function”, y denotes the vector of health care services, s denotes a random state of health whose subjective probability distribution obeys the probability law $\pi(s|\mathbf{A})$, \mathbf{A} is a vector of individual attributes, both observed and unobserved and d_j is the chosen health plan. We assume that $H(\cdot|\mathbf{A}, d_j)$ is increasing in y .

The consumer chooses health plans from a finite set of discrete and mutually exclusive alternatives, denoted d_j ($j = 1, \dots, J$), with insurance premia (P_1, \dots, P_J) . The consumer solves the allocation problem

$$\max_{\{d_j, C, \mathbf{y}\}} EU_j = \int U(C, H(\mathbf{y}, s|\mathbf{A}, d_j)) d\pi(s|\mathbf{A}) \quad (1)$$

subject to the constraint $C(s) + p_j u(s) = Y - P_j$, where p_j is the vector of net real out-of-pocket prices of health care services resulting from the choice of insurance plan j and Y is income.

Optimization by the dynamic programming approach involves two steps. First, conditional on choice of plan j , and each possible state of nature s , optimizing values of $C(s)$ and $y(s)$ are obtained. These solutions are substituted back into the utility function to derive optimizing values of the conditional expected (indirect) utility associated with each choice d_j , denoted EV_j , ($j = 1, \dots, J$). The d_j

which maximizes EV_j is chosen.

How does self-selection arise in this context? Optimizing individuals, possessing knowledge of their own health attributes, proclivities, and economic constraints, select plans accordingly. Self-perceived healthy individuals, expecting lower demand for future health care may choose low-cost plans with fewer choices than their less healthy counterparts. Others may have preferences for certain modes of care, e.g., office-based care from their family physician, and hence may choose plans with generous benefits in those dimensions. Therefore these attributes which partly determine the individual's choice of health plans also affect their expected utilization of services. Thus the presence of common factors induces correlation between the two sets of choices. A failure to control for such correlation is expected to distort econometric estimates of the impact of health plans on utilization. We regard the direction and magnitude of such distortion as an empirical issue.

The statistical issue of individual selection into plans can also arise from the economic behavior of health plans (Frank, et al., 2000; Cao and McGuire, 2002). Health plans that are offered by employers are often paid mostly through capitation or fixed payments. In such cases, profit-oriented health plans have an incentive to distort the quality of services they offer to attract profitable and deter unprofitable enrollees. For example, if demand for treatment of expensive chronic conditions is better anticipated and more unevenly distributed in a population than demand for less expensive acute care, then the health plan has an incentive to distort the mix of its care away from chronic care and towards acute illness in order to deter the high risks and attract the low risks. Frank, et al. (2000) show how the incentives to distort services depend in a relatively straightforward way on means and correlations among predicted values of health care services in a population. In an empirical analysis, they find that if people are assumed to know a few of their own relevant characteristics (age, sex and prior spending) selection incentives can be quite severe.

3. Econometric Model

Let EV_j^* denote the (latent) indirect utility associated with the j^{th} insurance plan, with $j = 0, 1, 2$ corresponding to plans without gatekeepers (*WOG*), managed care plans with gatekeepers but are not HMOs (*MCG*), and HMOs (*HMO*), respectively. Let d_j be binary variables representing the observed choices. Following convention, we treat *WOG* ($j = 0$) as the baseline choice. The indirect utility or propensity to select insurance plan j is formulated as

$$EV_{ji}^* = \mathbf{z}_i' \boldsymbol{\alpha}_j + \delta_j l_{ji} + \eta_{ji}. \quad (2)$$

where \mathbf{z}_i denotes exogenous covariates, $\boldsymbol{\alpha}_j$ the associated parameters and η_{ji} are random error terms. The l_{ji} are independently distributed random variables which we treat as latent factors. These latent factors are composites of variables such as unobserved components of individual and family health history, attitudes towards health risks, lifestyle choices etc., that influence individual perceptions of health events. The δ_j are factor loadings: parameters associated with the latent factors. The transformation from the latent variable formulation to the observed choices is via a distribution function \mathbf{g} that describes a multinomial choice model such that

$$\Pr(d_{ji} = 1 | \mathbf{z}_i, l_{ji}) = \mathbf{g}(\mathbf{z}_i' \boldsymbol{\alpha}_j + \delta_j l_{ji} + \eta_{ji}), \quad j = 0, 1, 2. \quad (3)$$

In this paper, we assume that \mathbf{g} has a mixed multinomial logit structure (MMNL) defined as

$$\Pr[d_{ji} = 1 | \mathbf{z}_i, l_{ji}] = \frac{\exp(\mathbf{z}_i' \boldsymbol{\alpha}_j + \delta_j l_{ji} + \eta_{ji})}{\sum_{k=0}^J \exp(\mathbf{z}_i' \boldsymbol{\alpha}_k + \delta_k l_{ki} + \eta_{ki})} \quad (4)$$

with the normalization restrictions $\alpha_0 = 0$ and $\delta_0 = 0$. This model is derived from maximization of utility function with random components, where the “sources of randomness in the utility function are unobserved variations in tastes and in the attributes of alternatives, and errors of perception and optimization by the consumer” (McFadden, 1980, p. S15).

Let y_i^* denote the value of the latent variable underlying the observed values of utilization, y_i . The outcome or utilization equation is formulated as

$$y_i^* = \mathbf{x}_i' \boldsymbol{\beta} + \gamma_1 d_{1i} + \gamma_2 d_{2i} + \sum_j \lambda_j l_{ji} + \varepsilon_i \quad (5)$$

where \mathbf{x}_i is a set of exogenous covariates and β , γ_1 , and γ_2 are parameters associated with the exogenous covariates and insurance dummy variables. The error term is partitioned into ε_i , an independently distributed random error, and l_{ji} which denotes unobserved characteristics common to individual i 's choice of insurance plan of type j and health services utilization of that individual. The λ_j are factor loadings. The transformation from y_i^* given in (5) to the observed random variable y_i is through an appropriate distribution function \mathbf{f} such that

$$\Pr(Y_i = y_i | \mathbf{x}_i, l_{ji}) = \mathbf{f}(\mathbf{x}_i' \boldsymbol{\beta} + \gamma_1 d_{1i} + \gamma_2 d_{2i} + \sum_j \lambda_j l_{ji} + \varepsilon_i). \quad (6)$$

Measures of utilization of curative health care services are reported as counts, $y_i = 0, 1, 2, \dots$, so we specify \mathbf{f} as the negative binomial-2 density (Cameron and Trivedi, 1998),

$$f(y_i | \mu_i) = \frac{\Gamma(y_i + \psi)}{\Gamma(\psi) \Gamma(y_i + 1)} \left(\frac{\psi}{\mu_i + \psi} \right)^\psi \left(\frac{\mu_i}{\mu_i + \psi} \right)^{y_i}, \quad (7)$$

where the conditional mean parameter $\mu_i = \exp(\mathbf{x}_i' \boldsymbol{\beta} + \gamma_1 d_{1i} + \gamma_2 d_{2i} + \sum_j \lambda_j l_{ji})$ denotes the mean component of utilization and $\psi \equiv 1/\alpha, (\alpha > 0)$ is an overdispersion parameter in the conditional variance $\mu_i (1 + \psi \mu_i)$. Utilization of preventive health care services are measured using a dichotomous variable denoting whether care was received during a period that roughly follows current medically recommended accepted standards of care. The length of the period for each measure of care is described in Section 4 below. We specify f as the normal distribution for such outcomes, i.e., a Probit model.

Because the latent factors l_{ji} enter both the insurance choice (3) and the utilization (6) equations, they capture the individual-specific (or idiosyncratic) factors that induce self-selection into insurance plans through unobservables on utilization of health care services. Observe also that such a specification explicitly

incorporates heterogeneity in the response of utilization to insurance plan. Idiosyncratic factors that induce variations in insurance coverage also directly impact on utilization. From a statistical perspective, $(\delta_j l_{ji} + \eta_{ji})$ and $(\sum_j \lambda_j l_{ji} + \varepsilon_i)$ are correlated even though $(\eta_{ji}, \varepsilon_i)$ are an uncorrelated pair, and potentially generate selection bias.

Under these assumptions, the joint distribution of selection and outcome variables, conditional on the common latent factors, can be written as

$$\begin{aligned} \Pr(Y_i = y_i, d_{ji} = 1 | \mathbf{x}_i, \mathbf{z}_i, l_{ji}) &= \mathbf{g}(\mathbf{z}_i' \alpha_j + \delta_j l_{ji} + \eta_{ji}) \\ &\times \mathbf{f}(\mathbf{x}_i' \beta + \gamma_1 d_{1i} + \gamma_2 d_{2i} + \sum_j \lambda_j l_{ji} + \varepsilon_i). \end{aligned} \quad (8)$$

A major problem in estimation arises because the l_{ji} are unknown. Although the l_{ji} are unknown, we assume that the distribution of l_{ji} , \mathbf{h}_j , is known. We assume that \mathbf{h}_j are standard normal densities, but we also conduct robustness checks with non-normal densities. The zero mean assumption is without loss of generality and fixed variance is needed because the variance of the latent factors cannot be separately identified. A normalization is required on either λ_j or δ_j because the variance is not identified. We assume $\delta_j = 1$ for each j and estimate values of λ_j . In addition, since $\delta_0 = 0$ and $\alpha_0 = 0$ are required for normalization in the multinomial logit model, we assume $l_{0i} = 0$ without loss of generality. Hence, l_{1i} and l_{2i} are interpreted as factors favoring *MCG* and *HMO* to *WOG*.

Under the assumptions stated above, the joint likelihood function of selection and outcome variables, conditional on the common latent factors, can be written as

$$L(y_i, d_{ji} | \mathbf{x}_i, \mathbf{z}_i, l_{ji}) = \prod_{i=1}^N \Pr(Y_i = y_i, d_{ji} = 1 | \mathbf{x}_i, \mathbf{z}_i, l_{ji}). \quad (9)$$

The likelihood conditional only on observables is obtained by integrating out the latent variables l_{ji} and the integration part can be performed numerically

rather than analytically thus:

$$\begin{aligned}
L(y_i, d_{ji} | \mathbf{x}_i, \mathbf{z}_i, l_{ji}) &= \prod_{i=1}^N \int \Pr(Y_i = y_i, d_{ji} = 1 | \mathbf{x}_i, \mathbf{z}_i, l_{ji}) \mathbf{h}_j(l_{ji}) dl_{ji} \\
&\equiv \mathbb{E} [L(y_i, d_{ji} | \mathbf{x}_i, \mathbf{z}_i, l_{ji})] \\
&\approx \prod_{i=1}^N \frac{1}{S} \sum_{s=1}^S \Pr(Y_i = y_i, d_{ji} = 1 | \mathbf{x}_i, \mathbf{z}_i, l_{jis}),
\end{aligned} \tag{10}$$

where the second line follows by the definition of expectations, and the third line is a numerical approximation of the integral obtained by averaging each term in the likelihood over S draws of l_{ji} from its assumed parametric distribution.

Our modeling strategy focuses on the utilization equation as the structural equation with a causal interpretation. The plan choice equations are also structural in the sense that they embody choice behavior. But their primary role is to yield good estimates of choice probabilities so we do not attempt a structural interpretation of the parameters of the plan choice equations. This estimation approach seems appropriate as we do not have information on insurance plan premia and related variables that play a key role in structural modeling of insurance choice. Hence, some may regard our plan choice equations as being of the reduced form variety.

The maximum simulated likelihood (MSL) estimator involves maximizing the simulated likelihood (Gouriéroux and Monfort, 1996). Provided that S is sufficiently large, the precise number being a function of N , the maximization of the simulated likelihood is equivalent to maximizing the likelihood. The literature recommends that S should increase faster than \sqrt{N} , but this does not give explicit guidance in choosing S . In univariate cases, a small number of random draws S is sufficient to reduce the simulation error to acceptable levels. However, it is well known that many more draws are required in multidimensional cases to achieve a similar level of accuracy. Increasing the number of simulation draws is simple in principle but computationally costly. Instead, we use “intelligent” systematic draws rather than random draws to speed up convergence of the expectation. The use of Halton sequences is one such quasi-Monte Carlo method

(Bhat, 2001; Train, 2002). Bhat (2001) found that 100 Halton draws provided more precise results for the mixed logit than 1000 random draws. We describe Halton sequences in Appendix 1.

In the work reported here we have used $S = 2000$ based on Halton draws. Note that this is a considerably larger number than has been used in many empirical studies that use the MSL method, e.g. Munkin and Trivedi (1999). Our experience with simulation based methods indicates that the number of simulations required for good approximation is considerably larger in models with endogenous regressors than in the models without such a complication. Further, the adequacy of any choice of S also depends upon how good the initial starting values are. In our case the starting values were obtained by initially estimating the plan choice equations and the outcome equation under the restriction of exogenous choice dummies.

We maximize the simulated likelihood using a quasi-Newton algorithm requiring only first derivatives. Post-convergence the variance of the MSL estimates is obtained using the usual sandwich formula for the covariance matrix. Information matrix and outer product formulae are inappropriate because they do not take into account uncertainty due to simulation chatter (McFadden and Train, 2000).

Marginal effects of covariates on the outcomes are also calculated by simulation. We calculate marginal effects for dummy variables as discrete changes and for continuous variables using derivatives. We calculate marginal effects for hypothetical values of all other covariates, e.g., at means or medians of the covariates. Note that each of these calculations requires averaging over simulated draws of the latent factors. Standard errors of the marginal effects are calculated using a Monte Carlo technique using 500 replications.

3.1. Testing and Interpreting Selection Effects

Single equation estimation of the outcome equation, under the assumption that the managed care variables d_{ji} (or treatment variables) are exogenous, would be an appropriate methodology if the treatments were randomly assigned, but this

is obviously not the case here. Provided that the model is correctly specified, the MSL estimates of (γ_1, γ_2) have the same interpretation as that under random assignment of treatments; $\gamma_1 > 0$ means that the treatment leads on average to an increase in utilization relative to the untreated state (here *WOG*).

The selection effect is measured by the factor loadings $(\lambda_{MCG}, \lambda_{HMO})$. If $\lambda_{HMO} < 0$, then the unobserved heterogeneity which makes an individual more likely to select the HMO causes that individual to have a utilization level for a service that is on average lower than that under randomized assignment. Because we expect healthier individuals to be more likely to choose plans with restrictions, we interpret such an effect as evidence of favorable selection. When $\lambda_j > 0$ favorable selection is indicated for analogous reasons.

Computation of the marginal impact of an insurance plan, and its sampling variance, in a nonlinear model is considerably more complex than in a linear model. This quantity essentially measures the difference between the additional utilization of an individual who is randomly assigned a particular treatment, i.e., *HMO* or *MCG*, and that of the typical individual with the benchmark health plan, *WOG*. Conceptually, for insurance plan j , the average treatment effect,

$$ATE(\mathbf{x}) \equiv E[y|\mathbf{x}, d_j = 1] - E[y|\mathbf{x}, d_j = 0]. \quad (11)$$

measures the effect on utilization of a specific health care service of randomly assigning the health plan j to an average individual with characteristics \mathbf{x} .

The idiosyncratic component of the change measured by

$$\Delta_s = E[u_1|\mathbf{x}, d_j = 1] - E[u_0|\mathbf{x}, d_j = 0] \quad (12)$$

should be negligible under a random assignment of treatment, but will be nonzero if exogeneity is assumed when the plans are self-selected. Our joint model of insurance plans and utilization statistically corrects for self-selection so $\Delta_s \approx 0$. In other words, when we calculate

$$\widehat{ATE}(\mathbf{x}_i, \widehat{\boldsymbol{\theta}}) \equiv E[y_i|\widehat{\boldsymbol{\theta}}, \mathbf{x}_i, d_{ji} = 1] - E[y_i|\widehat{\boldsymbol{\theta}}, \mathbf{x}_i, d_{ji} = 0], \quad (13)$$

where $E[y_i|\hat{\boldsymbol{\theta}}, \mathbf{x}_i, d_{ji}]$ is the conditional mean function evaluated at the MSL parameter estimates $\hat{\boldsymbol{\theta}}$, we expect $\Delta_s \approx 0$ and $\widehat{ATE}(\mathbf{x}_i, \hat{\boldsymbol{\theta}})$ to be an estimate of the causal treatment effect.

One oft-reported measure of the estimated ATE is obtained by evaluating the sample average of the estimated conditional means for all sample value of covariates \mathbf{x}_i , i.e.,

$$\widehat{ATE}(\mathbf{x}, \hat{\boldsymbol{\theta}}) = \frac{1}{N} \sum_{i=1}^N \left(E[y_i|\hat{\boldsymbol{\theta}}, \mathbf{x}_i, d_{ji} = 1] - E[y_i|\hat{\boldsymbol{\theta}}, \mathbf{x}_i, d_{ji} = 0] \right). \quad (14)$$

Although straightforward in principle, the standard errors of $\widehat{ATE}(\mathbf{x}, \hat{\boldsymbol{\theta}})$ were very computationally time consuming because they required Monte Carlo replications in addition to simulations within each replication. Therefore, instead we report effects and their standard errors at the sample average of the covariates, and at the median values of the covariates, the sample averages of covariates within specific subgroups of empirical interest (e.g. blacks, females, those with serious chronic conditions) and on the treated group (often referred to as the average treatment effect on the treated).

3.2. Identification of Causal Parameters

Issues of model identification arise due to the introduction of endogenous insurance dummies. The identification of the causal parameters through nonlinear functional forms is feasible in principle, but for more robust identification the traditional approach is through nontrivial exclusion restrictions or instrumental variables. Therefore, we need to find variables in the dataset that are correlated with the choice of health plan but are, conditional on exogenous variables in the outcome equation, uncorrelated with the outcomes. We use employment characteristics as identifying instruments. These variables are whether the individual is employed, whether the individual is self employed or works in the government sector, whether the individual belongs to a union, number of employees in the firm and whether it is in multiple locations, as well as indicators for industry sectors.

Johnson and Crystal (2000) and Olson (2002) also use employment characteristics as instruments in similar contexts.

We recognize that employment and access to health insurance may be jointly determined (Gruber, 2000), so we eliminate individuals who do not have private insurance coverage from our sample. Our instruments are plausibly assumed to affect the choice of type of health plan (conditional on having a health plan) but not utilization except indirectly through health plan choice. Consequently, our results should be treated as identifying the causal effects of plan type *conditional* on having insurance coverage. In addition, because it may be possible to argue that employment status and self-employment status among those employed may be jointly determined with the desire to have access to a particular type of health plan, we also estimate models for a subsample of those who are employed and a subsample of only those who are employees of firms using only firm characteristics as instruments.

Our model of health plan choice assumes that each individual has each type of plan available to choose from. We recognize that some individuals work for employers who do not offer any choice of health plans. On the other hand, some of these individuals have the option of being covered under plans from other members in the family. It is also generally possible to purchase all types of health insurance plans on the individual market at some finite price and within the budget set of the individual (although these prices are quite high). As long as the probability of choosing a health plan that is outside the set offered by the employer is not zero, our econometric model is not inconsistent with the data. Therefore, for econometric simplicity we assume that all individuals can choose among all types of plans.

4. Data

In this study, we use data from the 1996 Medical Expenditure Panel Survey (MEPS). MEPS has wide scope and contains excellent information on demographic characteristics, health status, employment status and earnings, and a

wide variety of measures of health care utilization. In our study we focus on the subsample of non-elderly adults (ages 18 to 64) who have some form of private health insurance. We eliminate individuals who are covered by Medicaid or other public insurance plans and Medicare enrollees (both elderly and disabled) because we wish to focus on the gatekeeper role of HMO's and other managed care plans among persons who make such choices in the private market for health insurance.

The effect of gatekeeping in managed care is captured via two dummy variables, enrollment in an *HMO* (47%) and enrollment in other managed care plans with gatekeepers denoted *MCG* (8.2%). The remainder are in plans that do not have gatekeeper restrictions to care denoted *WOG*. Enrollment status is measured at the first round of the survey in 1996.

Our empirical analysis covers five curative and five preventive measures of health care utilization. The first set of curative utilization variables are frequencies of visits to different types of providers: to an MD in an office setting, to a non-MD medical professional in an office setting, to a hospital, to the emergency room and to a hospital outpatient clinic ($N = 8129$). The second set of preventive care services are binary variables: whether blood pressure ($N = 7952$) and cholesterol checks ($N = 7717$) were received in the last two years, whether a flu shot was taken in the last year ($N = 7948$), and for females only, whether a pap smear ($N = 4082$) and a mammogram ($N = 2105$) was received in the last year. Summary statistics are presented in Table 1. Relative to individuals in *WOG* plans, those in *MCG* have significantly higher doctor visits while those in *HMO* plans have significantly higher outpatient utilization. Persons in *MCG* and *HMO* plans are more likely to have received blood pressure checks and women in these plans are more likely to have received pap smears. Finally, those in HMOs are more likely to have their cholesterol checked than individuals in *WOG* plans.

Our choice of explanatory variables for the utilization and insurance choice equations is similar to that in Dowd, et al. (1991), Ettner (1997) and Goldman, et al. (1995). Socioeconomic characteristics include age, for which we have explored

polynomial and linear spline specifications, gender, ethnicity, marital status, education, family size, location of residence, and personal income. Health characteristics include self-perceived health status, which we decompose into four dummy variables from the 5 point scale representing very good, good, fair and poor health (excellent health is the excluded category), the existence of a functional limitation and the number of chronic conditions. The determinants of insurance choice include all the socioeconomic and health characteristics that determine health care utilization. In addition, we include employment characteristics.

Descriptions and summary statistics of demographic, employment and health status control variables stratified by insurance plan choice are presented in Table 2. Individuals enrolled in *WOG* plans have significantly different demographic characteristics than those enrolled in *HMO* plans and, although to a lesser extent, those who are enrolled in *MCG* plans. Employment characteristics are different too. Most noticeable are differences in firm size, measured both by number of employees (*firmsize*) and whether the firm is in one or more locations (*multlocation*). There are no statistically significant differences in observed health status measures. Although others have found differences in observed health status across insurance plan types (see, e.g., Mello, et al., 2002), these studies are about other populations and/or include the uninsured. Our sample consists largely of individuals who receive health insurance as an employment benefit.

5. Results

In this section we discuss the results from ten jointly estimated models. We begin by discussing the insurance choice equations. Then we discuss utilization, grouped into curative and preventive categories.

5.1. Insurance Choice

The estimates of the MMNL insurance equations from each of the ten models are very similar because they are all estimates for the same choices of type of health plans with the same sets of covariates. So we present and discuss estimates from

only one of these models, that from the joint model of insurance and visits to the doctor. Marginal effects from this model are presented in Table 3. We find that older and rural individuals are more likely to choose *WOG* plans and less likely to choose *MCG* and *HMO* plans. Women and minorities are less likely to enroll in *WOG* plans and more likely to choose *HMO* plans. There are substantial regional differences as well. Health status indicators, educational attainment and income are generally not significant. These are reasonable results given that estimates are for insured only, most of whom obtain insurance from their employers (or from the employers of someone in the household). The insignificance of the health status variables in the choice equations suggests that for this particular population we do not have evidence of favorable selection on the basis of *observed* health status into HMOs. However, it is still possible that there is favorable selection on the basis of *unobserved* health status.

The insurance choice equations contain eight employment related variables that are excluded from the utilization equation. *HMO* enrollment probability is significantly, positively, and robustly related to being employed at a large firm (*firmsize*) with multiple locations (*multlocation*), and negatively related to *selfemployed*. On the other hand, employment sector and occupation are not significant. These instruments are tested for joint significance in the MMNL using the likelihood ratio (LR) statistic and are statistically significant in each case. For example, the LR test statistic is 125 for the sample used to estimate the model for doctor visits. This is large relative to the conventional $\chi^2(16)$ critical values and confirm that the instruments are statistically suitable identifiers.

5.2. Curative Health Care Services

Table 4 provides the estimated coefficients on the insurance dummy variables and the factor loadings associated with the latent factors for curative health care services.¹ The coefficient of the *HMO* dummy variable is positive and highly sig-

¹The full set of parameter estimates for the outcome equations is reported in Table 1 of Appendix 2. The estimated coefficients are of plausible sign and significance.

nificant for three measures: *Doctor*, *Outpatient* and *ER*. Thus, after correcting for self-selection, HMOs which have strong gatekeeper restrictions, encourage the use of curative health care in a number of potentially cost-effective dimensions. Unfortunately, they also tend to promote the use of emergency room services, perhaps because they are treated as the primary mode of “after hours” care. The factor loading coefficient λ_{HMO} is estimated to be negative and highly significant in three equations (*Doctor*, *Outpatient* and *ER*) but λ_{MCG} is typically not. The interpretation of the significantly negative factor loading coefficient is that the unobserved factors that increase the probability of being enrolled in an HMO also lead to lower utilization relative to that of the randomly assigned HMO enrollee. This means that there is favorable selection on unobservables into the HMO plans.

Table 5 presents treatment effects of *HMO* and associated standard errors for a variety of hypothetical individuals. For comparison, we have calculated the effects from our joint model which account for endogeneity of plan-type and from single-equation models which do not account for endogeneity. Given the imprecise nature of the estimates on *MCG* coefficients, we do not report treatment effects with respect to *MCG*. The hypothetical individuals we consider have the average characteristics of the entire sample, of black individuals, of non-black individuals and of males and females. We also calculate treatment effects at the average characteristics of the sample of individuals with no chronic conditions and those with one or more chronic conditions. Finally, we calculate treatment effects at the median characteristics of individuals in the sample and the average characteristics of those actually enrolled in HMOs. When endogeneity of plan-type is not accounted for, doctor visits are the only curative care with a statistically significant treatment effect. However, once self-selection is accounted for, doctor visits, outpatient visits and emergency room visits all have statistically significant treatment effects. For the individual with average characteristics and controlling for self-selection, those in HMOs are predicted to have 2.6 more doctor visits, 0.5 more outpatient visits and 0.13 more emergency room visits.

In each case, the treatment impacts controlling for self-selection are much larger than the corresponding treatment effects assuming exogeneity.

The magnitudes of treatment effects obtained for the “average individual” are very similar to those obtained for characteristics set at the sample averages of individuals who are actually enrolled in HMOs, i.e., the “average treated individual”. But the treatment effects for individuals who have median characteristics are substantially smaller, although statistically significant, than individuals with average characteristics. This demonstrates that the effect of being in an HMO differs substantially across individuals in the sample. The treatment effects are uniformly smaller for the average black individual as compared to the average non-black, for the average male as compared to the average female (except in the case of emergency room visits for which the treatment effects are very close) and for the average individual with chronic conditions as compared to the average individual with no chronic conditions. These results collectively suggest that different groups of individuals react differently to the incentives and restrictions on care implied by gatekeeper models of health care provision.

5.3. Preventive Health Care Services

Table 6 provides the estimated coefficients on the insurance dummy variables and the factor loadings associated with the latent factors for preventive health care services.² The coefficient of the *HMO* dummy variable is positive and highly significant for three measures: *Bloodpressure*, *Cholesterol* and *Flushot*. In addition, it is positive and marginally significant for *Mammogram*. In general, after correcting for self-selection, HMOs which have strong gatekeeper restrictions, encourage the use of preventive health care. For *MCG* enrollees the evidence is weak and statistically insignificant, except in the case of *Mammogram* where it is negative and it is statistically significant. The factor loading coefficient λ_{HMO} is estimated to be negative and highly significant for *Bloodpressure*, *Cholesterol*

²The full set of parameter estimates for the outcome equations is reported in Table 2 of Appendix 2. The estimated coefficients are of plausible sign and significance.

and *Flushot*, but λ_{MCG} is typically not. Once again, the interpretation of the significantly negative factor loading coefficient is that the unobserved factors that increase the probability of being enrolled in an HMO also lead to lower likelihoods of receiving preventive care relative to that of the randomly assigned HMO enrollee.

treatment effects of *HMO*, calculated for a variety of hypothetical individuals, are reported in Table 7. For comparison, we have calculated the effects from our joint model which account for endogeneity of plan-type and from single-equation models which do not account for endogeneity. Because the outcome variables are binary, these treatment effects are the changes in probabilities of receiving the preventive health care services. Once again, the hypothetical individuals we consider have the average characteristics of the entire sample, of black individuals, of non-black individuals and of males and females, of sick and healthy, of those actually enrolled in HMOs and a hypothetical individual with median values of characteristics. Individuals enrolled in *HMO* plans (relative to *WOG*) are 10, 28, 21 and 20 percentage points more likely to receive blood pressure checks, cholesterol exams, flu shots and mammograms, respectively. The effect of *HMO* on papsmear tests is not significant. These estimated plan impact effects on probabilities of service are between 2 and 10 times larger as compared to estimates assuming exogeneity of *HMO* status. Moreover, although there are significant and substantial *HMO* effects on *flushot* and *mammogram* when the endogeneity of health-plan type is considered, these effects are small and insignificant in the single-equation models that do not account for endogeneity.

The effect sizes obtained for the “average individual” are, once again, very similar to those obtained for characteristics set at the sample averages of individuals who are actually enrolled in HMOs, i.e., the “average treated individual”. For preventive care, however, there is no clear relationship between the treatment effects calculated for the median individual as compared to the treatment effects calculated for the average individual. The effect of *HMO* enrollment is smaller for the average black individual as compared to the average non-black for blood-

pressure, cholesterol and flusht. The average male has a greater *HMO* effect than an average female with respect to blood pressure and cholesterol checks but the relative effect size is reversed for flu shots. A similar pattern is observed when one compares effects sizes for the healthy as compared to the sick. Generally, the effect sizes across hypothetical females is very similar for papsmear and mammogram.

5.4. Robustness Checks

Estimates of complex econometric models can be sensitive to choices of samples and covariates, distributional assumptions and parametric functional forms. In order to inform on such issues, our estimated models are subjected to six robustness checks, two involving variations in the sample coverage, two more for the unobserved heterogeneity assumption, and the last two in respect of the estimation method used. These results are summarized in Tables 8 and 9 which report parameter estimates and marginal effects of *HMO* for curative and preventive care respectively.

Our first robustness check examines the sensitivity of estimated parameters to variations in sample coverage. Recall that our results identify the causal effects of plan type conditional on having insurance coverage. But we have argued that employment status and self-employment status among those employed may be jointly determined with the desire to have access to a particular type of health plan. Therefore, we estimate models for a subsample of those who are employed and a subsample of only those who are employees of firms using only firm characteristics as instruments. Tables 8 and 9 show how the estimated marginal impact of *HMO* changes if we reestimate our models after excluding first the unemployed and then both the unemployed and the self-employed from the full sample. For example, the qualitative impact on doctor visits is to reduce the estimate without much change in the standard error. Relative to the full sample, the marginal impact of *HMO* drops to 2.53 and 2.24 visits compared to 2.65 in the full sample. When the same exercise is carried out for outpatient visits, hospital discharges,

emergency room visits, the results regarding the *HMO* impact are, after allowing for the expected sampling variation, very similar to those for the full sample. For all five preventive measures estimated impact retains the same sign and roughly the same size as in the full original sample.

Our second check involves the impact of using alternative distributional assumptions for the latent factors. In place of normality we assume that the latent factors are drawn from beta distributions centered at zero with unit variance. We consider two cases. In the first, the parameters of the beta density are chosen to have skewness equal to 0.5 and in the second, the selected parameters give a skewness of -0.5. The impact of applying the MSL methodology to these new specifications on the conclusions about the point estimates of the impact of *HMO* is fairly small. Broadly, the count outcomes show relatively greater sensitivity than the binary outcomes.

Because, our estimation procedure is of full information variety, and such procedures may be sensitive to model misspecification, our third robustness check involves using two simpler “limited information” alternatives based on the instrumental variables method. We mimic the linear two stage least squares approach by estimating linear probability equations for *HMO* choices and then substituting the fitted probabilities in place of the *HMO* dummy in the negative binomial model for counted measures for curative outcomes and the binary measures for the preventive outcomes. These results are shown against the label “models with fitted plan choice”. However, because of the nonlinearity of the outcome equations this procedure does not in general yield consistent estimates, although such procedures are sometimes employed for convenience (Dubin and McFadden, 1984; Johnson and Crystal, 2000). We also use an alternative approach which is justified for a linear outcome equation. For curative utilization the outcome equation linearized by taking logarithms³ In the case of preventive utilization variables the outcome is a dummy variable, and hence the outcome model is of a linear probability equation. We use a linear instrumental variables (LIV) procedure in

³A small positive value is added to the count to avoid definitional problems for zero counts.

which the instrument set consists of all the exogenous covariates in the outcome and insurance choice equations. The results in Tables 8 and 9 are sign-wise consistent with those from MSL estimation. The direct impact of *HMO* on counted outcomes is in the same direction as MSL estimates, but generally estimated with considerably less precision, especially when fitted probabilities are used. For binary outcomes, the two sets of results are much closer in terms of point estimates, but the LIV estimates are very imprecise. For example, the results for the blood pressure checks show that the marginal effect of *HMO* is nearly twice as large under MSL methods than under IV assumptions, and both are larger than under exogeneity assumptions, but with large standard errors.

In summary, the results of the robustness exercise provide strong support for the use of a structural latent variable framework to obtain efficient estimates of the key parameters. There is, however, one robustness check that we have not implemented. We have used the MMNL model with the independence from irrelevant alternatives (IIA) property. It is desirable that we relax this strong assumption that will fail to hold if, for example, the plan choices are not distinct alternatives. The multinomial probit (MNP) model is a leading flexible alternative to the MMNL. However, its use in the present context is not feasible because the identification of the covariance structure in the MNP model requires alternative-variant exclusion restrictions. When alternative-specific covariates such as prices are available, as is usually the case in models of transportation choice, the identifying information exists in a usable form. However, here all data are individual-specific and generation of alternative-specific covariates can be done only somewhat arbitrarily (see, for example, Lechner, 2002). Finally, note that even with alternative specific covariates identification of the MNP can be quite fragile (Keane, 1992).

6. Concluding Remarks

We have used computer intensive simulation-based methods to jointly model the choice of health insurance plans and health care utilization respecting the multino-

mial nature of insurance choice, the discreteness of utilization, and the possibility of self-selection into insurance plans. Contrary to much existing econometric research on health care utilization that assumes exogeneity of insurance plans, we find significant evidence of selection bias. We find evidence of favorable selection into HMOs, i.e., individuals who are more likely to enroll in HMOs are likely to utilize less curative care and less likely to receive preventive care, *ceteris paribus*. We show that HMOs, which have strong gatekeeping features, encourage the use of physician and outpatient hospital curative care and the use of preventive services. On the other hand, HMOs appear to encourage the use of emergency room care as well. We speculate that the reason for this observation is that HMOs may use the emergency room as the preferred mode of after-hours care, but an investigation of possible reasons is beyond the scope of this research. Finally, we do not observe any effects of gatekeeping on the use of hospital care. We have attempted to use a number of exogenously defined subsamples to limit the sample to those who might be at risk for hospitalizations in an attempt to identify significant effects, but these searches have been unsuccessful. Therefore, we conclude that insurance plans with gatekeepers are unable to modify hospital use behavior in any significant way.

Our study has some notable limitations. First, beyond the gatekeeping feature, we do not have plan characteristics such as benefit features and premiums. If other plan characteristics are correlated with gatekeeping, it is possible that demand side incentives rather than supply side ones are driving the results. Second, the estimates of coefficients related to other plans with gatekeepers are imprecisely estimated. It is possible that this is simply due to the fact that only 8 percent of individuals are in such plans. However, it may also be due to measurement error in determination of *MCG* status. Finally, we have estimated ten separate equations instead of a joint model with ten outcomes. As the outcomes are correlated, a joint model would yield more efficient estimates. We deemed such a high-dimensional model computationally infeasible with currently available technology.

Issues of self-selection in non-normal, nonlinear contexts arise in many important problems in health economics. Many involve the appearance of more than one endogenous treatment dummy variable in an equation for a discrete or censored outcome. The approach developed here, which uses a latent factor structure to model endogeneity and maximum simulated likelihood for estimation, can be extended quite generally. Thus our methods and experience should be of use in other research areas as well. However, because our computational methods are very time intensive, further research is needed to investigate other promising computational methodologies, especially Bayesian approaches, that would make it feasible to efficiently handle models and samples larger than those used in this article.

References

- Bhat, C.R. (2001) “Quasi-random Maximum Simulated Likelihood Estimation of the Mixed Multinomial Logit Model”, *Transportation Research: Part B*, 35, 677-693.
- Cameron, A.C., P.K. Trivedi, F. Milne and J. Piggott (1988), “A Microeconomic Model of the Demand for Health Care and Health Insurance in Australia”, *Review of Economic Studies*, 55, 85-106.
- Cameron, A.C., and P.K. Trivedi (1998), *Regression Analysis of Count Data*, Econometric Society Monograph, New York: Cambridge University Press.
- Cao, Z. and T.G. McGuire (2002), “Service-Level Selection by HMOs in Medicare”, *Working Paper*.
- Christensen, S. and J. Shinogle (1997), “Effects of Supplemental Coverage on Use of Services by Medicare Enrollees”, *Health Care Financing Review*, 19, 5-17.
- Cutler, D.M. and R.J. Zeckhauser (2000), “The Anatomy of Health Insurance”, in Culyer, A.J. and J.P. Newhouse, Ed., in *Handbook of Health Economics*, Vol. 1A, 563-644.
- Dowd, B., R. Feldman, S. Cassou, and M. Finch (1991), “Health Plan Choice and the Utilization of Health Care Services”, *Review of Economics and Statistics*, 73, 85-93.
- Dubin, J.A. and D. McFadden (1984), “An Econometric Analysis of Residential Electric Appliance Holdings and Consumption”, *Econometrica*, 52, 345-362.
- Ettner, S.L. (1997), “Adverse Selection and the Purchase of Medigap Insurance by the Elderly”, *Journal of Health Economics*, 16, 543-62.
- Frank, R.G., J. Glazer and T.G. McGuire (2000), “Measuring Adverse Selection in Managed Health Care”, *Journal of Health Economics*, 19, 829-854.

- Gabel, J. (1997), "Ten Ways HMOs Have Changed During the 1990s", *Health Affairs*, 16, 134-144.
- Gilleskie, D. (1998), "Dynamic Stochastic Model of Medical Care Use and Work Absence", *Econometrica*, 66, 1-45.
- Glied, S. (2000), "Managed Care", Chapter 13 in Culyer, A.J. and J.P. Newhouse, Editors, *Handbook of Health Economics*, Vol 1A, 707-745.
- Goldman, D.P. (1995), "Managed Care as a Public Cost-Containment Mechanism", *Rand Journal of Economics*, 26, 277-95
- Goldman, D.P., S.D. Hosek, L.S. Dixon, E.M. Sloss (1995), "The Effects of Benefit Design and Managed Care on Health Care Costs", *Journal of Health Economics*, 14, 401-18.
- Gouriéroux, C. and A. Monfort (1996), *Simulation Based Econometrics Methods*, New York: Oxford University Press.
- Gruber, J. (2000), "Health Insurance and Labor Market", *Handbooks of Health Economics Vol. 1A*, Edited by Culyer, A. J. and J.P. Newhouse. Amsterdam: Elsevier.
- Heckman, J.J., J. Smith and N. Clements (1997), "Making the Most of Programme Evaluations and Social Experiments: Accounting for Heterogeneity in Programme Impacts", *The Review of Economic Studies*, 64, 487-535.
- Heckman, J.J., and E. Vytlačil (2001), "Policy-Relevant Treatment Effects", *American Economic Review Papers and Proceedings*, 91, 107-111.
- Johnson, R.W. and S. Crystal (2000), "Uninsured Status and Out-of-pocket Costs at Midlife", *Health Services Research*, 35, 911-932.
- Keane, M.P. (1992), "A Note on Identification in the Multinomial Probit Model", *Journal of Business and Economic Statistics*, 10, 193-200.

- Kemper, P., H.T. Tu, J.D. Reschovsky and E. Schaefer (2001), “Insurance Product Design and Its Effects: Trade-offs Along the Managed Care Continuum”, Working Paper, Center for Studying Health System Change.
- Lechner, M. (2002), “Program Heterogeneity and Propensity Score Matching: An Application to the Evaluation of Active Labor Market Policies”, *Review of Economics and Statistics*, 84, 205-220.
- McFadden, D. (1980), “Econometric Models for Probabilistic Choice among Products”, *Journal of Business*, 53, S13-S29.
- McFadden, D. and K. Train (2000), “Mixed MNL Models of Discrete Response”, *Journal of Applied Econometrics*, 15, 447-470.
- Mello, M.M., S.C. Stearns, E.C. Norton (2002), “Do Medicare HMOs Still Reduce Health Services Use After Controlling for Selection Bias?”, *Health Economics*, 11, 323-340.
- Miller, R.H. and H.S. Luft (1994), “Managed Care Plan Performance Since 1980”, *Journal of American Medical Association*, 271, 1512-1519.
- Munkin M.K. and P.K. Trivedi (1999), “Simulated Maximum Likelihood Estimation of Multivariate Mixed-Poisson Regression Models With Application”, *Econometrics Journal*, 2, 29-48.
- Newhouse, J.P. and the Insurance Experiment Group (1993), *Free for all? Lessons from the RAND Health Insurance Experiment*, Cambridge and London: Harvard University Press.
- Olson, C. (2002), “Do Workers Accept Lower Wages in Exchange for Health Benefits”, *Journal of Labor Economics*, 20, S91-S114.
- Tu, H.T., P. Kemper, and H.J. Wong (1999), “Do HMOs Make a Difference? Use of Health Services”, *Inquiry*, 36, 400-410.

Train, K. (2002), *Discrete Choice Methods with Simulation*, New York: Cambridge University Press.

Table 1: Descriptive Statistics for Utilization

Variable	Definition	N	HMO	MCG	WOG
Curative utilization			47%	8.2%	44.8%
Doctor	number of visits to a physician in an office setting	8129	3.279	3.499*	3.066
Nondoctor	number of non-physician visits in an office setting	8129	1.378	1.694	1.423
Hospital	number of hospital discharges	8129	0.076	0.070	0.082
ER	number of emergency room visits	8129	0.141	0.124	0.138
Outpatient	number of hospital outpatient visits	8129	0.386*	0.368	0.545
Preventive utilization					
Bloodpressure	=1 if blood pressure was checked in last two years	7952	0.923*	0.924*	0.887
Cholesterol	=1 if cholesterol was checked in last two years	7717	0.630*	0.595	0.564
Flu shot	=1 if flu shot was received in the last year	7948	0.198	0.185	0.198
Pap smear	=1 if pap smear test was received in the last year	4082	0.679*	0.694*	0.617
Mammogram	=1 if mammogram was received in the last year	2105	0.537	0.531	0.521

Note: * indicates that the estimate is significantly different from the base case (*WOG*) at the 5 percent level.

Table 2: Descriptive Statistics for Explanatory Variables

Variable	Definition	HMO	MCG	WOG
Demographic characteristics				
familysize	family size	3.080	2.993	3.042
age	age/10	3.944*	3.937*	4.091
education	years of school	13.474	13.572	13.395
income	income/1000	39.014*	37.514	37.272
female	=1 if female	0.534*	0.513	0.505
black	=1 if black	0.122*	0.118*	0.084
hispanic	=1 if hispanic	0.158*	0.125*	0.093
married	=1 if married	0.674	0.715	0.683
northeast	=1 if north east	0.212*	0.183	0.194
midwest	=1 if midwest	0.200*	0.244*	0.295
south	=1 if south	0.321*	0.367	0.352
msa	=1 if metropolitan statistical area	0.871*	0.906*	0.697
Employment characteristics				
employed	=1 if employed	0.888*	0.887	0.863
selfemployed	=1 if self employed	0.083*	0.082*	0.130
firmsize	firm size/10	14.675*	14.481*	10.581
multlocation	=1 if multiple locations	0.594*	0.629*	0.499
union	=1 if union	0.148*	0.124	0.129
govtjob	=1 if government job	0.183*	0.151	0.161
blue	=1 if blue collar	0.223	0.204	0.216
service	=1 if service	0.356	0.392*	0.346
miscellaneous	=1 if miscellaneous industry	0.086	0.086	0.076
Health status				
verygood	=1 if very good health	0.356	0.349	0.365
good	=1 if good health	0.239	0.249	0.227
fair	=1 if fair health	0.061	0.052	0.056
poor	=1 if poor health	0.012	0.010	0.014
chronic	number of chronic conditions	0.541	0.517	0.535
physicallim	=1 if physical limitation	0.056	0.064	0.059

Note: * indicates that the estimate is significantly different from the base case (*WOG*) at the 5 percent level.

Table 3: Marginal Effects in MMNL Insurance Plan Choice Model

Variable	Pr(WOG)		Pr(MCG)		Pr(HMO)	
	Marg.	St. err.	Marg.	St. err.	Marg.	St. err.
familysize	0.009*	0.004	-0.006*	0.002	-0.003	0.004
age	0.034*	0.006	-0.006*	0.003	-0.028*	0.006
married	-0.050*	0.015	0.029*	0.007	0.021	0.014
northeast	0.102*	0.018	-0.009	0.009	-0.093*	0.017
midwest	0.175*	0.018	0.002	0.010	-0.177*	0.016
south	0.127*	0.016	0.006	0.009	-0.133*	0.015
msa	-0.241*	0.014	0.057*	0.006	0.185*	0.014
income	-3e-4	2e-4	-2e-4	1e-4	0.001*	2e-4
female	-0.035*	0.012	-0.005	0.006	0.040*	0.012
black	-0.085*	0.018	0.006	0.011	0.079*	0.019
hispanic	-0.082*	0.018	-0.004	0.009	0.085*	0.018
education	4e-4	0.003	0.000	0.002	-0.001	0.003
employed	0.014	0.022	-0.001	0.013	-0.013	0.023
selfemployed	0.050*	0.023	-1e-4	0.013	-0.050*	0.022
firmsize	-0.002*	4e-4	3e-4	2e-4	0.002*	4e-4
multlocation	-0.057*	0.015	0.024*	0.008	0.033*	0.015
union	0.010	0.018	-0.012	0.009	0.002	0.017
govtjob	-0.015	0.019	-0.016	0.009	0.031	0.019
blue	-0.009	0.019	-0.008	0.010	0.017	0.019
service	-2e-4	0.016	0.002	0.009	-0.001	0.015
physicallim	-0.024	0.026	0.017	0.016	0.007	0.027
chronic	-0.009	0.008	-0.001	0.004	0.010	0.008
verygood	-0.002	0.015	-0.004	0.007	0.006	0.014
good	-0.033*	0.016	0.007	0.008	0.026	0.016
fair	-0.020	0.027	-0.007	0.014	0.027	0.028
poor	0.026	0.053	-0.014	0.027	-0.012	0.052

Note: * indicates that the parameter estimate is significantly different from zero at the 5 percent level.

Table 4: Insurance and factor loading parameters: curative health care services

	Doctor		Nondoctor		Outpatient		Hospital		ER	
	Coef.	St. err.	Coef.	St. err.	Coef.	St. err.	Coef.	St. err.	Coef.	St. err.
MCG	0.126	0.158	0.216	0.158	0.635*	0.173	0.378	0.568	0.296	0.325
HMO	0.906*	0.051	0.047	0.086	1.396*	0.107	-0.547	0.455	0.928*	0.149
λ_{MCG}	0.136	0.169	0.001	0.001	-0.878*	0.088	-0.460	0.569	-0.348	0.325
λ_{HMO}	-0.934*	0.047	-0.001	0.002	-1.686*	0.078	0.648	0.532	-1.004*	0.156

Note: * indicates that the parameter estimate is significantly different from zero at the 5 percent level.

Table 5: Marginal effects of HMO: curative health care services

	Doctor		Nondoctor		Outpatient		Hospital		ER	
	Marg.	St. Err.	Marg.	St. Err.	Marg.	St. Err.	Marg.	St. Err.	Marg.	St. Err.
accounting for endogeneity of health plan choice										
mean	2.649*	0.197	0.047	0.085	0.531*	0.075	-0.035	0.096	0.127*	0.031
median	1.898*	0.151	0.038	0.070	0.222*	0.041	-0.033	0.105	0.056*	0.016
black	1.895*	0.164	0.041	0.062	0.277*	0.087	-0.021	0.094	0.114*	0.031
non black	3.065*	0.220	0.105	0.101	0.679*	0.090	-0.034	0.091	0.154*	0.031
male	1.955*	0.142	0.063	0.060	0.405*	0.058	-0.025	0.066	0.156*	0.031
female	3.905*	0.279	0.119	0.119	0.771*	0.103	-0.045	0.136	0.150*	0.031
chronic>0	4.998*	0.364	0.185	0.186	1.137*	0.159	-0.048	0.140	0.197*	0.040
chronic=0	1.952*	0.137	0.056	0.053	0.367*	0.049	-0.027	0.076	0.131*	0.026
in HMO's	2.812*	0.179	0.088	0.083	0.492*	0.063	-0.035	0.108	0.148*	0.030
assuming exogeneity of health plan choice										
	Marg.	St. Err.	Marg.	St. Err.	Marg.	St. Err.	Marg.	St. Err.	Marg.	St. Err.
mean	0.247*	0.090	0.043	0.083	-0.035	0.044	-0.004	0.006	0.013	0.009
median	0.194*	0.070	0.034	0.067	-0.020	0.027	-0.004	0.006	0.007	0.005
black	0.194*	0.070	0.031	0.061	-0.040	0.060	-0.003	0.005	0.014	0.009
non black	0.254*	0.093	0.045	0.087	-0.034	0.043	-0.004	0.006	0.013	0.009
male	0.174*	0.064	0.030	0.060	-0.029	0.037	-0.003	0.004	0.013	0.009
female	0.341*	0.124	0.059	0.114	-0.042	0.053	-0.005	0.008	0.014	0.009
chronic>0	0.433*	0.157	0.093	0.181	-0.066	0.084	-0.006	0.009	0.018	0.012
chronic=0	0.174*	0.064	0.026	0.051	-0.023	0.030	-0.003	0.005	0.011	0.008
in HMO's	0.246*	0.089	0.042	0.081	-0.034	0.044	-0.004	0.006	0.013	0.009

Note: * indicates that the parameter estimate is significantly different from zero at the 5 percent level.

Table 6: Insurance and factor loading parameter: preventive health care services

	Bloodpressure		Cholesterol		Flushot		Papsmear		Mammogram	
	Coef.	St. err.	Coef.	St. err.	Coef.	St. err.	Coef.	St. err.	Coef.	St. err.
MCG	0.365	0.471	0.234	0.321	0.206	0.377	-0.282	0.385	-1.129*	0.482
HMO	1.032*	0.403	1.141*	0.277	1.532*	0.209	0.610	0.540	1.050	0.622
λ_{MCG}	-0.077	0.502	-0.148	0.341	-0.193	0.392	0.508	0.424	1.388*	0.447
λ_{HMO}	-0.927*	0.430	-1.120*	0.302	-1.750*	0.229	-0.570	0.615	-1.177	0.727

Note: * indicates that the parameter estimate is significantly different from zero at the 5 percent level.

Table 7: Marginal effects of HMO: preventive health care services

	Bloodpressure		Cholesterol		Flushot		Papsmear		Mammogram	
	Marg.	St. Err.	Marg.	St. Err.	Marg.	St. Err.	Marg.	St. Err.	Marg.	St. Err.
accounting for endogeneity of health plan choice										
mean	0.102*	0.026	0.283*	0.035	0.210*	0.015	0.177	0.119	0.199*	0.091
median	0.139*	0.036	0.294*	0.036	0.195*	0.016	0.167	0.116	0.199*	0.091
black	0.076*	0.029	0.260*	0.033	0.185*	0.014	0.280*	0.116	0.241*	0.090
non black	0.093*	0.026	0.296*	0.035	0.213*	0.015	0.281*	0.119	0.241*	0.091
male	0.123*	0.036	0.297*	0.035	0.204*	0.015	—	—	—	—
female	0.065*	0.019	0.289*	0.034	0.217*	0.015	—	—	—	—
chronic>0	0.048*	0.016	0.256*	0.030	0.243*	0.017	0.279*	0.118	0.240*	0.090
chronic=0	0.124*	0.036	0.303*	0.036	0.188*	0.014	0.283*	0.120	0.241*	0.091
in HMO's	0.090*	0.027	0.294*	0.035	0.204*	0.015	0.281*	0.118	0.241*	0.091
assuming exogeneity of health plan choice										
	Marg.	St. Err.	Marg.	St. Err.	Marg.	St. Err.	Marg.	St. Err.	Marg.	St. Err.
mean	0.026*	0.005	0.058*	0.012	0.013	0.009	0.045*	0.016	0.023	0.024
median	0.033*	0.008	0.061*	0.013	0.013	0.009	0.042*	0.015	0.023	0.024
black	0.024*	0.006	0.051*	0.011	0.011	0.008	0.043*	0.015	0.022	0.024
non black	0.026*	0.005	0.059*	0.012	0.013	0.009	0.045*	0.016	0.023	0.024
male	0.040*	0.008	0.060*	0.012	0.012	0.009	—	—	—	—
female	0.016*	0.003	0.057*	0.012	0.014	0.010	—	—	—	—
chronic>0	0.011*	0.003	0.050*	0.010	0.016	0.011	0.044*	0.016	0.023	0.024
chronic=0	0.039*	0.008	0.061*	0.013	0.011	0.008	0.045*	0.016	0.023	0.024
in HMO's	0.026*	0.006	0.058*	0.012	0.013	0.009	0.045*	0.016	0.023	0.024

Note: * indicates that the parameter estimate is significantly different from zero at the 5 percent level.

Table 8: HMO Effects on Curative Care: Alternative Models and Samples

Model	N	Coeff.	St. Err.	Marg.	St. Err.
doctor visits					
Sample without unemployed	7127	0.914*	0.061	2.531*	0.293
Sample without unemployed and self employed	6285	0.803*	0.065	2.246*	0.201
Beta distributed latent factors skewness = 0.5	8129	0.906*	0.055	2.649*	0.212
Beta distributed latent factors skewness = -0.5	8129	0.906*	0.052	2.682*	0.202
Linear instrumental variables	8129	1.038	1.284	1.038	1.284
Models with fitted insurance plan choice	8129	0.549	0.325	1.414	0.838
non doctor visits					
Sample without unemployed	7127	0.043	0.094	0.040	0.090
Sample without unemployed and self employed	6285	0.036	0.100	0.035	0.092
Beta distributed latent factors skewness = 0.5	8129	0.047	0.086	0.047	0.085
Beta distributed latent factors skewness = -0.5	8129	0.047	0.086	0.047	0.085
Linear instrumental variables	8129	2.653	1.731	2.653	1.731
Models with fitted insurance plan choice	8129	1.086	0.733	1.072	0.724
outpatient visits					
Sample without unemployed	7127	1.411*	0.126	0.556*	0.098
Sample without unemployed and self employed	6285	1.397*	0.128	0.526*	0.089
Beta distributed latent factors skewness = 0.5	8129	1.457*	0.117	0.598*	0.103
Beta distributed latent factors skewness = -0.5	8129	1.385*	0.107	0.628*	0.099
Linear instrumental variables	8129	0.950	0.789	0.95	0.789
Models with fitted insurance plan choice	8129	2.681*	0.908	0.816*	0.277
hospital discharges					
Sample without unemployed	7127	-0.552*	0.210	-0.032*	0.015
Sample without unemployed and self employed	6285	-0.532*	0.262	-0.032	0.021
Beta distributed latent factors skewness = 0.5	8129	-0.511	0.384	-0.032	0.059
Beta distributed latent factors skewness = -0.5	8129	-0.544*	0.267	-0.035	0.022
Linear instrumental variables	8129	-0.027	0.090	-0.027	0.090
Models with fitted insurance plan choice	8129	-0.531	1.133	-0.032	0.069
emergency room visits					
Sample without unemployed	7127	0.935*	0.162	0.127*	0.032
Sample without unemployed and self employed	6285	0.928*	0.241	0.135	0.119
Beta distributed latent factors skewness = 0.5	8129	0.738*	0.166	0.097*	0.027
Beta distributed latent factors skewness = -0.5	8129	0.726*	0.153	0.096*	0.026
Linear instrumental variables	8129	0.205	0.126	0.205	0.126
Models with fitted insurance plan choice	8129	1.421	0.854	0.170	0.102

Note: * indicates that the parameter estimate is significantly different from zero at the 5 percent level.

Table 9: HMO Effects on Preventive Care: Alternative Models and Samples

Model	N	Coeff.	St. Err.	Marg.	St. Err.
blood pressure check					
Sample without unemployed	6969	1.034	0.592	0.104*	0.049
Sample without unemployed and self employed	6137	1.045*	0.318	0.107*	0.019
Beta distributed latent factors skewness = 0.5	7952	2.131*	1.066	0.135*	0.040
Beta distributed latent factors skewness = -0.5	7952	1.029*	0.377	0.102*	0.024
Linear instrumental variables	7952	0.138	0.077	0.180*	0.079
Models with fitted insurance plan choice	7952	1.397*	0.490	0.174*	0.063
cholesterol check					
Sample without unemployed	6763	1.173*	0.220	0.289*	0.026
Sample without unemployed and self employed	5959	1.107*	0.210	0.279*	0.027
Beta distributed latent factors skewness = 0.5	7717	1.173*	0.309	0.285*	0.041
Beta distributed latent factors skewness = -0.5	7717	1.162*	0.291	0.285*	0.036
Linear instrumental variables	7717	0.494*	0.137	0.494*	0.137
Models with fitted insurance plan choice	7717	1.507*	0.356	0.577*	0.136
flu shot					
Sample without unemployed	6971	1.557*	0.192	0.210*	0.013
Sample without unemployed and self employed	6145	1.591*	0.212	0.211*	0.013
Beta distributed latent factors skewness = 0.5	7948	1.532*	0.234	0.214*	0.016
Beta distributed latent factors skewness = -0.5	7948	1.530*	0.220	0.213*	0.015
Linear instrumental variables	7948	0.539*	0.126	0.539*	0.126
Models with fitted insurance plan choice	7948	2.084*	0.386	0.542*	0.100
mammogram					
Sample without unemployed	1675	0.937	0.817	0.192	0.119
Sample without unemployed and self employed	1490	1.427*	0.668	0.273*	0.081
Beta distributed latent factors skewness = 0.5	2105	1.032	0.803	0.197	0.113
Beta distributed latent factors skewness = -0.5	2105	1.033	0.602	0.196*	0.089
Linear instrumental variables	2105	0.149	0.161	0.149	0.161
Models with fitted insurance plan choice	2105	0.537	0.415	0.214	0.165
papsmear					
Sample without unemployed	3357	0.613	0.780	0.178	0.140
Sample without unemployed and self employed	3040	0.647	0.835	0.188	0.159
Beta distributed latent factors skewness = 0.5	4082	0.609	0.610	0.175	0.130
Beta distributed latent factors skewness = -0.5	4082	0.609	0.629	0.175	0.128
Linear instrumental variables	4082	0.136	0.153	0.136	0.153
Models with fitted insurance plan choice	4082	0.355	0.427	0.131	0.157

Note: * indicates that the parameter estimate is significantly different from zero at the 5 percent level.

Appendix 1: Description of Halton Sequences

Increasing the number of simulation draws to reduce the simulation error to acceptable levels is simple in principle but computationally costly. In our case, computational times were prohibitively high when sufficient numbers of pseudo-random draws were used. In numerical analysis a literature has recently emerged that attempts to use intelligent, systematic draws rather than random draws to speed up convergence of the required expectations. The quasi-Monte Carlo method is similar to the Monte Carlo method but instead of using S pseudo-random points, it uses non-random points within the domain of integration. The use of Halton sequences is one such quasi-Monte Carlo method introduced by Bhat (2001) in the context of simulation-based estimation of mixed multinomial models. Halton sequences have two desirable properties vis-a-vis pseudo-random draws. First, they are designed to give fairly even coverage over the domain of the mixing distribution. With more evenly spread draws for each observation, the simulated probabilities vary less over observations, relative to those calculated with random draws. Second, with Halton sequences, the draws for one observation tend to fill in the spaces left empty by the previous observations. The simulated probabilities are, therefore, negatively correlated over observations. This negative correlation reduces the variance in the simulated likelihood function. Under suitable regularity conditions, the integration error using pseudo-random sequences is in the order of N^{-1} as compared to pseudo-random sequences where the convergence rate is $N^{-1/2}$ (Bhat, 2001).

Halton sequences are best described by example. Consider the prime number 2. Its Halton sequence is constructed as follows. Divide the unit interval (0,1) into 2 parts. The dividing point $1/2$ becomes the first element of the Halton sequence. Next divide each part into two more parts. The dividing points, $1/4$ and $3/4$ become the next two elements of the sequence. Divide each of the four parts into two parts each, and continue. Halton sequences on non-prime numbers are not unique because the Halton sequence for a non-prime number divides the unit space in the same way as each of the prime numbers that constitute the

non-prime.

In our model, we have two latent factors l_{1i} and l_{2i} that need to be integrated out. We begin by generating two Halton sequences based on the primes 2 and 3:

$$\begin{aligned}\xi_{1i} &= \{1/2 \quad 1/4 \quad 3/4 \quad 1/8 \quad 3/8 \quad \dots\} \\ \xi_{2i} &= \{1/3 \quad 2/3 \quad 1/9 \quad 2/9 \quad 4/9 \quad \dots\}\end{aligned}$$

The length of each sequence is determined by the number of observations N and the numbers of simulation draws S which we have chosen to be 2000. The early elements of Halton sequences with different primes have a tendency to be correlated with each other (see Train, 1999, for an example). Consequently, we begin by generating Halton sequences of length $N \times S + 20$ and discard the first twenty elements of each sequence. The required normally distributed quasi-random draws for l_{1i} and l_{2i} are generated by applying the inverse of the normal cumulative distribution function to the Halton sequences, i.e.

$$\begin{aligned}l_{1i} &= \Phi^{-1}(\xi_{1i}) \\ l_{2i} &= \Phi^{-1}(\xi_{2i}).\end{aligned}$$

The first group of S elements is assigned to the first observation in the sample, the next S elements to the second observation, and so on.

Appendix 2: Parameter estimates of Outcome Equations

Table 1: Curative Care

Variable	Doctor		Nondoctor		Outpatient		Hospital		ER	
	Coef.	St. err.	Coef.	St. err.	Coef.	St. err.	Coef.	St. err.	Coef.	St. err.
intercept	-1.726	0.147	-2.178	0.338	-5.881	0.391	-3.605	0.450	-1.816	0.342
familysize	-0.049	0.012	-0.099	0.032	-0.122	0.035	0.059	0.035	-0.008	0.030
age	0.061	0.015	0.07	0.040	0.268	0.041	-0.033	0.052	-0.184	0.039
education	0.057	0.007	0.097	0.018	0.041	0.019	-0.005	0.021	-0.046	0.017
income	0.001	0.001	0.002	0.001	-0.002	0.001	-2e-4	0.002	8e-5	0.002
female	0.615	0.032	0.561	0.090	0.536	0.090	0.581	0.105	-0.098	0.077
black	-0.353	0.052	-0.376	0.213	-0.37	0.188	-0.157	0.176	-0.103	0.125
hispanic	-0.163	0.052	-0.533	0.136	-0.402	0.153	0.162	0.15	0.006	0.121
married	0.112	0.038	0.037	0.098	0.282	0.108	0.187	0.122	-0.193	0.090
northeast	0.271	0.048	-0.08	0.119	0.882	0.138	0.129	0.152	0.223	0.121
midwest	0.173	0.047	-0.137	0.114	1.033	0.139	-0.129	0.160	0.522	0.116
south	0.209	0.044	-0.204	0.119	0.26	0.136	-0.041	0.139	0.215	0.111
msa	-0.051	0.042	0.019	0.106	-0.467	0.104	0.003	0.178	-0.389	0.094
verygood	0.192	0.038	0.283	0.104	0.292	0.114	-0.003	0.129	0.189	0.097
good	0.416	0.044	0.453	0.122	0.68	0.119	0.457	0.135	0.409	0.104
fair	0.797	0.066	0.349	0.177	1.318	0.178	1.245	0.172	1.089	0.142
poor	0.991	0.130	1.402	0.319	1.811	0.338	2.112	0.268	1.421	0.237
physicallim	0.260	0.062	1.164	0.249	0.486	0.169	0.319	0.170	-0.017	0.143
chronic	0.490	0.019	0.641	0.053	0.463	0.050	0.264	0.057	0.277	0.046
MCG	0.126	0.158	0.216	0.158	0.635	0.173	0.378	0.568	0.296	0.325
HMO	0.906	0.051	0.047	0.086	1.396	0.107	-0.547	0.455	0.928	0.149
α	0.278	0.060	6.163	0.223	1.139	0.199	1.692	0.877	0.574	0.283
λ_{MCG}	0.136	0.169	0.001	0.001	-0.878	0.088	-0.46	0.569	-0.348	0.325
λ_{HMO}	-0.934	0.047	-0.001	0.002	-1.686	0.078	0.648	0.532	-1.004	0.156
log likelihood	-24331		-16668		-12133		-9165		-10322	

Table 2: Preventive Care

Variable	Bloodpressure		Cholesterol		Flushot		Papsmear		Mammogram	
	Coef.	St. err.	Coef.	St. err.	Coef.	St. err.	Coef.	St. err.	Coef.	St. err.
intercept	-0.733	0.309	-3.837	0.575	-5.037	0.492	-0.272	0.361	-5.386	1.347
familysize	-0.082	0.026	-0.031	0.018	-0.110	0.030	-0.122	0.036	-0.117	0.061
age	0.039	0.030	0.432	0.063	0.463	0.051	-0.092	0.030	0.685	0.169
education	0.085	0.020	0.070	0.013	0.070	0.015	0.060	0.020	0.079	0.029
income	0.002	0.001	0.002	0.001	0.001	0.001	0.002	0.001	0.003	0.002
female	0.729	0.141	0.179	0.052	0.226	0.071	—	—	—	—
black	0.042	0.101	0.446	0.093	-0.454	0.128	0.192	0.091	0.211	0.197
hispanic	-0.077	0.094	0.180	0.074	-0.257	0.116	-0.032	0.097	-0.087	0.216
married	0.237	0.083	0.266	0.066	0.099	0.086	0.411	0.124	0.495	0.188
northeast	0.277	0.111	0.506	0.108	-0.080	0.103	-0.035	0.084	0.444	0.214
midwest	0.171	0.107	0.180	0.081	0.061	0.101	-0.031	0.094	0.306	0.212
south	0.193	0.100	0.367	0.089	0.102	0.095	0.033	0.088	0.106	0.172
verygood	0.120	0.068	0.079	0.055	0.067	0.082	-0.113	0.070	0.020	0.147
good	0.324	0.096	0.176	0.066	0.225	0.091	-0.035	0.075	0.092	0.161
fair	0.557	0.186	0.364	0.120	0.391	0.148	-0.166	0.127	0.302	0.255
poor	0.761	0.417	0.081	0.218	0.128	0.311	-0.382	0.245	0.450	0.485
physicallim	0.140	0.175	0.012	0.105	0.059	0.139	-0.198	0.118	-0.247	0.216
msa	-0.122	0.090	0.154	0.059	-0.396	0.100	0.117	0.098	0.280	0.168
chronic	0.579	0.129	0.377	0.060	0.246	0.047	0.097	0.039	0.083	0.071
MCG	0.365	0.471	0.234	0.321	0.206	0.377	-0.282	0.385	-1.129	0.482
HMO	1.032	0.403	1.141	0.277	1.532	0.209	0.610	0.540	1.050	0.622
λ_{MCG}	-0.077	0.502	-0.148	0.341	-0.193	0.392	0.508	0.424	1.388	0.447
λ_{HMO}	-0.927	0.430	-1.120	0.302	-1.750	0.229	-0.570	0.615	-1.177	0.727
log likelihood	-9085		-11295		-10571		-6085		-3174	